Army Technical Library Improvement Studies

Report No.18

MULTIPLE TEST OF ABC METHOD

PART III

Mathematical Model

Technical Information Office
Harry Diamond Laboratories
U.S. Army Material Command
Washington, D.C. 20438

May 1967

DA-1L013001A91A
AMCMS Code: 5016.11.84400
HDL Proj No. 01200

AD

TR-1334

MULTIPLE TEST OF ABC METHOD
PART III
Mathematical Model

by
Werner H. Menden

with contributions by
Bert Levy

May 1967

U.S. ARMY MATERIEL COMMAND
HARRY DIAMOND LABORATORIES
WASHINGTON, D.C. 20438

F O R E W O R D

The paper is a progress report on efforts to place a particular retrieval system on a mathematical foundation. The work is not complete in that some serious mathematical problems are still outstanding. However, it was felt desirable to record accomplishments for historical reasons and to invite critical reviews hopefully to gain insight leading to an improvement of the mathematical models for retrieval systems in general.

The authors have made an effort to minimize the formal mathematical exposition in the interest of documentalists with little or no mathematical training.

CONTENTS

ILLUSTRATIONS

ABSTRACT

The report suggests a method of constructing a mathematical model for the first test of the ABC Storage and Retrieval Systems and calculates 95-percent confidence intervals for relevance and recall values.

I. INTRODUCTION

This progress report on the evaluation of the first-generation ABC system is divided into five parts. In Part I, the subject is introduced, reference is made to reports previously published on the project, and the objectives are outlined. A few general remarks on the application of probabilistic models to the performance evaluation of retrieval tests are added.

In Part II, the test environment is described. Part III describes the statistical model on which the analysis of test data is based. In Part IV, point estimates for the relevance and recall parameters of the model are developed. In Part V, we deal with the accuracy and reliability of these estimates and develop confidence intervals for recall and relevance parameters of the ABC system.

In the appendices, various necessary formulas are derived.

A. Background

The ABC System (ref 1), developed by HDL* to effect efficient storage and retrieval of scientific and technical information, has been subjected to an extensive performance test. The first report in this series (ref 2) described the test program and setup as well as methodology, and the second report (ref 3) gave a preliminary statistical evaluation of test results. This third report of the series discusses a statistical model of the retrieval process, developed from a set of definitions and assumptions that were not refuted by experimental evidence. The model permits a rigorous analysis of the test data including the approximation of 95-percent confidence intervals for the relevance and recall performance of the system.

---

*The Army Research Office, Scientific and Technical Information Division, Washington, D. C., supports the development of the ABC System.

B. Why a Statistical, Probabilistic Model?

There is a growing literature* on the use of mathematical models for the description and analysis of information storage and retrieval systems. So far, the majority of the models suggested have been deterministic in that they allow one to predict the outcome of a retrieval experiment with certainty. These models make use of the fact that certain aspects of information systems** can be "mapped" into, or represented by, various abstract mathematical structures based on Boolean algebras, topology, lattice theory, set theory, and related descriptions.

Probabilistic models, on the other hand, have been used to study phenomena in information systems that obey statistical or probabilistic rather than deterministic laws. Among such phenomena, one might mention the number of monthly accessions of a collection, or the number of users in a given period of time. Obviously, the outcome of any particular retrieval run observed in our test (say four documents retrieved, three of which are relevant) also obeys probabilistic laws in a sense that it might not have been predicted with certainty.

One of the major purposes of this report is to study these probabilistic laws, and to specify them. The process of specification or choice of a particular statistical model requires acceptance of a few simplifying assumptions leading to the specific probabilistic equations of the model. Once the model equations have been formulated, estimators for the performance measures*** relevance and recall, can be derived and, subsequently, the accuracy of these estimators is determined in terms of confidence intervals.

II. TEST ENVIRONMENT

The test collection consisted of approximately 3650 documents (journal articles and reports) on solid-state circuits, devices, and applications. These covered the 1959-1964 period. The subject area was sufficiently small to permit comprehensive coverage of the open literature.

---

*An annotated bibliography is given at the end of the report.

**In particular, systems using computers for storage and/or retrieval.

***For definition, see Part II and III.

The main objective was to test the effectiveness of the ABC dictionary (a KWIC-type index of concise, informative, and indicative document descriptions in natural English). Two (one long, one short) dictionaries were used in the test. For the print-out of the long version, 250 terms were excluded from the permutation; in the short version, this list of terms was extended to include numerous noninformative terms; e.g., improvement and development.

A total of 136 questions were used for the test. These consisted of 100 questions based on source documents and 36 questions formulated by scientists with a broad overall knowledge of the contents of the collection.

Each retrieval run consisted in a single search performed by a test operator in one of the ABC dictionaries and resulted in a list of document descriptions. After the search was completed, the documents corresponding to the selected descriptions were analyzed by independent umpires for relevance to the question a hand, and the proportion of relevant documents could then be determined.

On the average, each document description was accessible (in other words, was permuted in the KWIC arrangement) under about four to five index terms.

The average number of relevant documents per question was determined to be 8.20 for the 36 general questions and 8.60 for the 100 source-document questions. (The procedure for assessing relevance is described in reference 3, page 6.)

Detailed information regarding test design and procedure are given in references 2 and 3.

III. DESCRIPTION OF THE STATISTICAL MODEL

A. Notation and Basic Definitions

The following letter symbols are used to denote the basic parameters and variables of the test.

N = number of documents available for retrieval in answer to a question

r = number of documents in the collection relevant to a given question (or a given set of questions, if explicitly stated)

x = number of documents retrieved in a retrieval run (by a retrieval operator in response to a given question) relevant to the question

y = number of documents retrieved in a retrieval run not relevant to the question at hand

n=x+y= number of documents retrieved in a retrieval run. Each retrieval run (or inquiry) consists of n trials; i.e., each document retrieved in answer to a question is a trial

x/n = relevance ratio for a retrieval run

x/r = recall ratio for a retrieval run

For the analysis of the test results, N and r are known parameters, and x, y, and n are random variables observed in the test.

B. Basic Assumptions and Model Equations

Probability Laws for x and y

For a retrieval run yielding n = x + y retrieved documents, it was assumed that x and y are independent and Poisson distributed;

therefore,

$$g(x,\mu) = \frac{e^{-\mu}\mu^{x}}{x!} \qquad (1)$$

$$g(y,\nu) = \frac{e^{-\nu}\nu^{y}}{y!} \qquad (2)$$

where $g(x,\mu)$ denotes the probability mass function (pmf) of x, and $g(y,\nu)$, the pmf of y. For a given x, $g(x,\mu)$is the probability of retrieving exactly x relevant documents, and μ is the expected value, or population mean, of x. A necessary condition for the independence of x and y is* that the retrieval of a relevant document at any particular trial is independent of the results at preceding trials. This condition was provided for to some extent by the test design in as much as relevance was judged by independent umpires after the retrieval run was completed.

The assumption that x and y are Poisson distributed was made because the Poisson distribution is often applied to the analysis of rare events involving chance processes, which the retrieval process is considered to be: the retrieval of a very few documents

*It can be shown that x and y are independent if and only if n is Poisson distributed

from a relatively large collection and whether a trial is relevant or not being a matter of chance. A more detailed presentation of rationale is presented in Appendix A.

χ^2 tests* were performed on numerous samples to check if the retrieval data fitted a Poisson distribution, and no sample tested gave evidence to refute the assumption.

Probability Law of x given n

Given that x and y are independently distributed according to equations 1 and 2, the pmf of x, given n documents have been retrieved, can be derived (Appendix B), and is found to be binomial:

$$f(x|n) = \binom{n}{x} p^x (1-p)^{n-x} \qquad (3)$$

The expected value** of x given n, which is denoted $E[x|n]$, is

$$E[x|n] = np$$

For any fixed value of x, say $x = k$, $f(k|n)$ is the probability of finding exactly k relevant documents among n documents retrieved. In a first approximation, we can assume $n \leq r$ for our tests***.

The parameter p can be interpreted as the probability that a document is relevant given that it is retrieved; therefore, p represents the precision (relevance) of the ABC system.

Probability Law for n

In a similar fashion, the pmf of n, say g(n), can be derived *** from (1) and (2) assuming the independence of x and y; n is also found to follow a Poisson distribution:

$$g(n) = \frac{e^{-\lambda}\lambda^n}{n!} \qquad (4)$$

with mean $E[n] = \lambda$; $\lambda = \mu + \nu$

---

*A common statistical method to test the goodness of fit of probability distribution.

**The expected value or "population mean" is the mean of the phenomenon being observed, weighted by its probability distribution.

***For the data analyzed, $n \leq r$ was found to hold for 98 percent of all runs.

****See Appendix B for derivation of g(n).

λ, the expected number of documents retrieved, can be regarded as a measure of the retrieval effort. Test results show λ to be dependent on r; therefore, we will always write $\lambda \equiv \lambda(r)$.

For any fixed value of n, say $n = m$, $g(m)$ specifies the probability that exactly m documents will be retrieved.

The validity of (4) was checked for various representative samples of the retrieval data, using again χ^2 methods to test the goodness of fit. Results show (4) to be a good approximation and also confirmed λ to be dependent on r.

So far, we have specified three basic probability laws, (1), (2), and (4) that allow us to predict, in terms of probabilities, the outcome of observations of our test variables x,y, and n alone. Furthermore, (3) allows us to predict a value of x, given that n has been observed before.

Now, we would like also to find the probability to observe any fixed value of x and a fixed value of n jointly.

According to statistical theory, the probability law for the joint distribution of x and n is obtained as the product of the pmf of n and the pmf of x, given n:

$$h(x,n,p) = f(x,p|n) \cdot g(n) \tag{5}$$

Using (3) and (4), we have

$$h(x,n,p) = \binom{n}{x} p^x (1-p)^{(n-x)} \frac{e^{-\lambda(r)} \lambda(r)^n}{n!} \tag{6}$$

The derivation of (6) completes our set of probability laws, which constitute the statistical model. The reader should recognize that the preceding section and the appendixes are of crucial importance in establishing correspondence between the physical world, i.e., the test data, and the mathematical model for this data. From here on, most results will be obtained by way of analytical derivation with little or no additional information about the physical world needed; therefore, the usefulness of the results derived will depend almost exclusively on the soundness of human judgment leading to the model assumptions and, consequently, to the formulation of the probabilistic laws. In the next part, we will be concerned with the derivation of formulas that can be used to estimate the parameters of the test, in particular, precision and recall, from observed sample values.

IV. ESTIMATION OF PRECISION AND RECALL

A. Estimation of Precision p and $\lambda(r)$, the Expected Number of Documents Retrieved

The method of maximum likelihood* was used to estimate the precision parameter p and the retrieval effort $\lambda(r)$ based on a series of k retrieval runs**.

Using the hat symbol ($\hat{}$) to denote estimators, we find

$$\hat{p} = \frac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{k} n_i} \tag{7}$$

Precision is estimated from a series of k observations not by averaging the relevance ratios (x_i/n_i), but by averaging x_i and n_i separately.

In similar fashion, an estimator for $\lambda(r)$ is derived:

$$\hat{\lambda}(r) = \frac{\sum_{i=1}^{k} n_i}{k} \tag{8}$$

The expected value of n for a single run is estimated from k observations of n as the arithmetic mean. Since experimental evidence showed λ to be dependent on r, the estimation of this parameter will be based on samples of n with constant r. In other words, the model appears to be applicable only to subsets of data. The same is true of recall ratio, which is discussed later.

Precision p was introduced in (3) as a basic parameter of the model. Therefore p should be independent of r; (7) can be used to check the validity of this statement. For this purpose, we ranked the 100 questions based on source documents into 24 groups corresponding to 24 values of r asserted for them. Then for each value of r, p was calculated using (7). Also, a value for $\hat{p}$ was obtained by extending the summations in (7) over all data observed in the test.*** The results are shown in figure 1.

---

*Described in many textbooks on statistics.

**See Appendix C for complete derivation.

***Including more than 1000 retrieval runs.

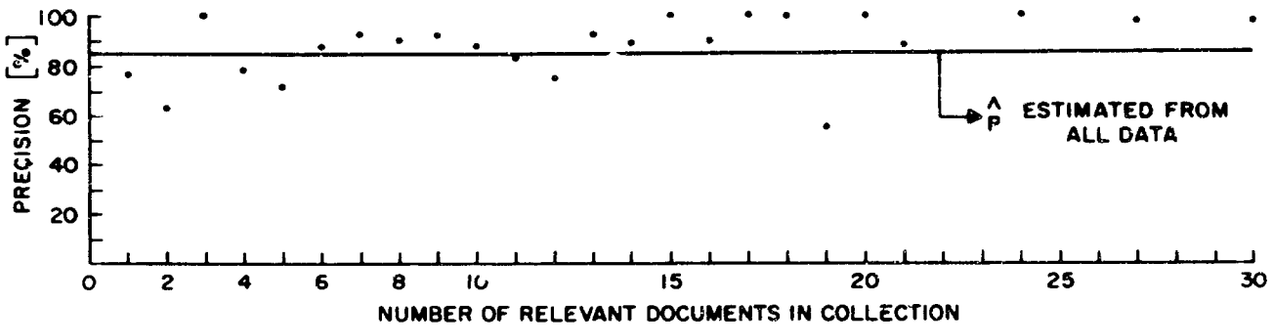

Figure 1. Precision versus number of relevant documents in collection. Data is from the 100 questions based on source documents. Precision was estimated by $\hat{p} = \sum_{i=1}^{k} x_i \Big/ \sum_{i=1}^{k} n_i$. Dots represent values of $\hat{p}$ calculated for each value of r from the corresponding subset of observations on (x,n). The solid line represents the overall systems estimate for p calculated from all data observed in the test: $\hat{p}$ = 85.7 percent.

Obviously, p is fairly independent of r in this first approximation, although a slight tendency for p to increase with r should not be overlooked.

B. Estimation of Recall

The precision parameter p appeared in our model as the parameter of the binomial probability law that governs the frequencies of observing any particular number of relevant documents, given that n documents have been retrieved. With recall, the situation is different. Since our basic model equations (1) and (2) do not contain r explicitly, the model cannot contain parameters that could be used to represent recall.* Therefore, we will introduce a recall parameter $\rho(r)$ by definition to be the expected value of the average recall ratio observed in the test.

Assume we observe the recall ratio x/r for a series of k retrieval runs with constant r:

$$x_1/r,\ x_2/r,\ \ldots,\ x_k/r$$

Now, let

$$\gamma_i = x_i/r;\ \bar{\gamma} = \frac{\sum_{i=1}^{k} \gamma_i}{k} = \frac{\sum_{i=1}^{k} x_i}{kr}$$

Then, we define a recall parameter $\rho(r)$ as the expected value of the average recall ratio $\bar{\gamma}$:

$$\rho(r) \equiv E[\bar{\gamma}] \qquad (9)$$

*The α model discussed in Appendix E of this report uses a different basic approach and contains a recall parameter α but no precision parameter.

Using the method of the moment generating function, $E[\bar{\gamma}]$ can be derived (Appendix D, Section 2) and is found to be:

$$E[\bar{\gamma}] = \frac{\lambda(r)p}{r} \tag{10}$$

$\rho(r)$ can be estimated using our previously derived estimates $\hat{\lambda}(r)$ and $\hat{p}$ to have

$$\hat{\rho}(r) = \frac{\hat{\lambda}(r)\ \hat{p}}{r} \tag{11}$$

Using the summations in (7) and (8), we obtain

$$\hat{\rho}(r) = \frac{1}{r} \quad \frac{\sum_{i=1}^{k} n_i}{k} \cdot \quad \frac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{k} n_i}$$

Since summations of n_i are over the same sets (r = constant) of data, they cancel each other and we obtain,

$$\hat{\rho}(r) = \frac{\sum_{i=1}^{k} x_i}{rk} = \bar{\gamma} \tag{12}$$

For a series of k retrieval runs with constant r, the recall parameter $\rho(r)$ can be estimated to a first approximation by the average recall ratio $\bar{\gamma}$ observed for the k runs. In figure 2, values of $\hat{\rho}(r)$ obtained from (12) have been plotted versus r. In the same diagram, a second set of values for $\hat{\rho}(r)$ is plotted, which was obtained from the general formula (11), using the overall systems estimate for $\hat{p}$ calculated from all test data. From the agreement between the two sets of values, it is evident the choice of $\hat{p}$—either from subsets of data with constant r or from all data—has little influence on the estimation of recall. The dependence of recall on r in figure 2 probably stems from the relatively low retrieval effort $\hat{\lambda}(r)$, which varied between 1 and 4 in the test. In the second report of this series (ref 3), it is shown that in the majority of cases, the observed average recall ratios were less than 10 percent below the optimum obtainable for a given n and r.

We have now derived estimates for precision and recall based on observations from the test. Also, an estimate for the retrieval effort $\lambda(r)$ was obtained. We found evidence for our assumption that the precision p is a true systems parameter of relevance; however, recall can be estimated only for subsets of retrieval data with constant r. In Part V, confidence intervals will be

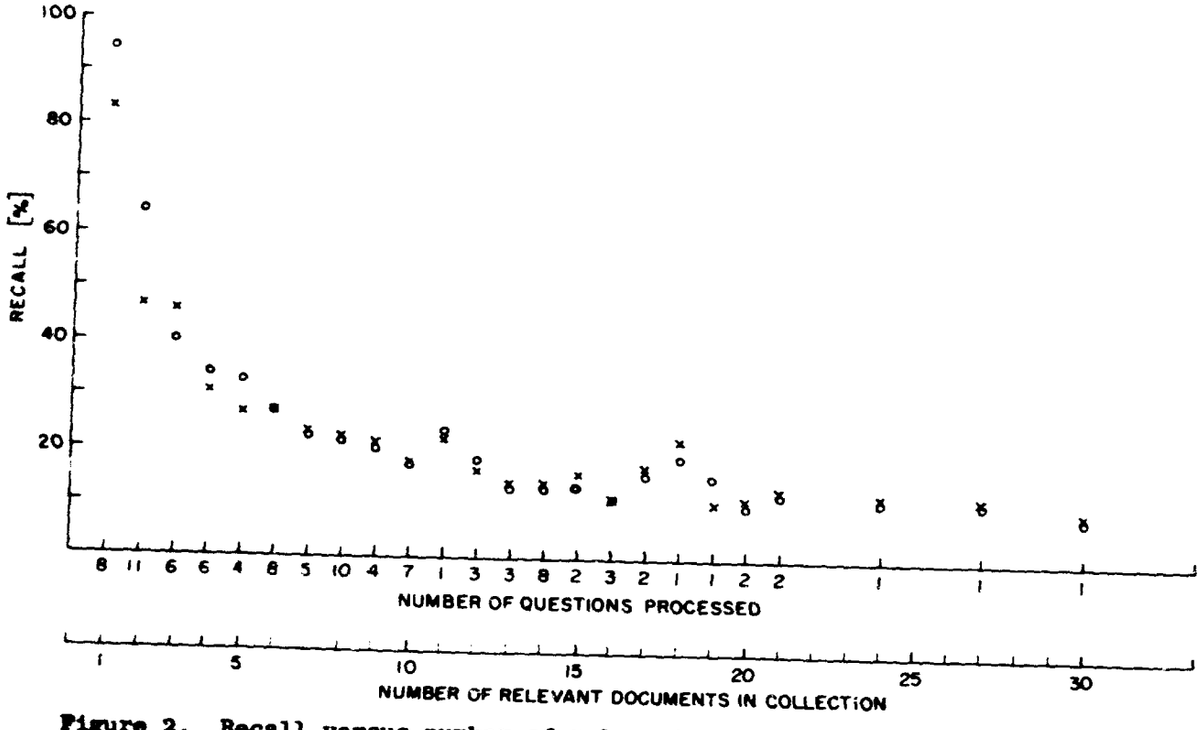

Figure 2. Recall versus number of relevant documents in collection. Data is from 100 questions based on source documents. The x indicate recall $\rho(r)$ estimated by the average recall ratio $\overline{\gamma}$. The o indicate recall $\rho(r)$ estimated by $\hat{\rho} = \hat{\lambda}(r)\,\hat{p}/r$ where $\hat{p}$ is calculated from all test data; $\hat{p} = 85.7$ percent.

established for both relevance and recall. The purpose is here to determine accuracy or reliability of our estimates derived so far.

V. DETERMINATION OF CONFIDENCE INTERVALS

While a detailed discussion of the significance of confidence intervals would be beyond the scope of this report, a few explanatory remarks seem appropriate for the nonstatistically minded reader. Assume a statistical parameter p to be estimated by calculating $\hat{p}$ for each of a large number of samples. Then for each sample, a confidence interval is determined so that p is contained within the intervals for a high proportion of samples, while it may fall outside the interval in a few cases. The small probability that p will fall outside the upper or lower limits of the intervals has to be preselected before the limits are calculated, and is called level of significance. For this report, the level of significance, often denoted α, is chosen to be 0.025. The value of 2α can be thought of as the probability of making an error in claiming that the calculated limits include p.

In symbolic notation, using $\hat{\underline{p}}$ to denote the lower limit and $\overline{\hat{p}}$ to denote the upper limit and P to denote the probability for the statement in brackets to be true

$$P[\hat{\underline{p}} < p < \overline{\hat{p}}] = 95 \text{ percent} \qquad (13)$$

Equation (13) is equivalent to the proposition: the true value of the parameter p will be contained in the limits as determined from samples in 95 percent of all cases, or, in roughly 95 percent of a limited number of samples analyzed. In a very loose sense confidence intervals may be likened to error bounds about the true value of the parameter.

If the preselected error probability 2α is small, the investigator may have a high degree of confidence in his assumption that the true value of the parameter estimated will fall within the limits of the error bands.

After these general remarks, we will give confidence bands about $\sigma(r)$ and p.

A. Confidence Intervals about Recall $\rho(r)$

In Appendix D, it is shown that $\bar{\nu}$ has a Poisson distribution with variable $nk\bar{\nu}$ and mean $k\lambda(r)p$

If we define:

$$\lambda'(r) = k\lambda(r)p, \qquad (14)$$

We can use statistical tables (ref 7) for confidence intervals about the mean of a Poisson distribution to find upper and lower limits for $\lambda'(r)$. The corresponding intervals about $\rho(r) = \lambda(r)p/r$ may then be obtained using the identity

$$\frac{\lambda(r)p}{r} \equiv \frac{\lambda'(r)}{kr} \tag{15}$$

and we can write

$$P\left[\frac{\underline{\lambda'(r)}}{kr} < \rho(r) < \frac{\overline{\lambda'(r)}}{kr}\right] = 0.95 \tag{16}$$

for a level of significance $\alpha = 0.025$.

In figure 3, confidence intervals about $\rho(r)$, estimated by $\hat{\rho}(r)$, are presented as vertical lines for each sample of data with fixed r. The number of retrieval runs is shown under each value of r, for r = 1 to 10. Some comment is needed for the sample with r = 1; here, the upper limit on $\rho(r)$ was found to be larger than 100 percent. The reason is that in 8 out of 59 valid retrieval runs involved, n exceeded r, which disagrees with our original assumption made for the model.

B. Confidence Intervals about Precision p

Since the distribution of the average relevance ratio for k runs is not known, intervals about p have been determined using the upper and lower limits for $\lambda'(r)$ tabulated (ref 7). By definition (14) we have

$$p = \frac{\lambda'(r)}{k\lambda(r)} \tag{17}$$

and using $\hat{\lambda}(r)$ to estimate the unknown parameter $\lambda(r)$ in (17), we get an estimator for p:

$$\hat{p} = \frac{\lambda'(r)}{k\hat{\lambda}(r)} \tag{18}$$

From (18), upper and lower limits on p can be obtained as

$$\underline{\hat{p}} = \frac{\underline{\lambda'(r)}}{k\hat{\lambda}(r)} \tag{19}$$

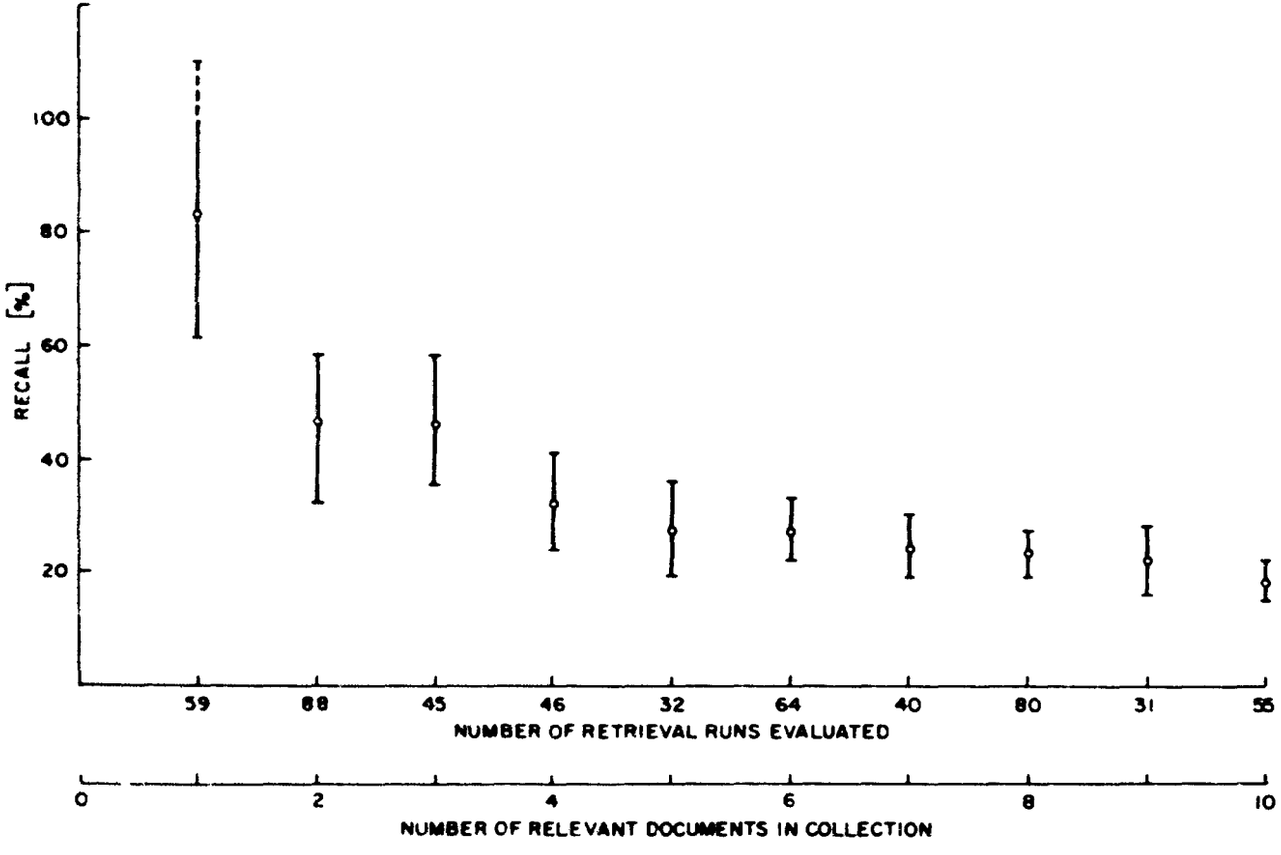

Figure 3. Confidence intervals about recall $\rho(r)$. $\rho(r)$ is estimated by $\hat{\rho}(r) = \bar{y}$. Level of significance $\alpha = 2.5$ percent.

and

$$\bar{\hat{p}} = \frac{\overline{\lambda'(r)}}{k\hat{\lambda}(r)} \qquad (20)$$

For several samples, upper limits on p determined from (20) exceeded 100 percent. This seeming inconsistency is, however, easily explained:

(1) the relevance as estimated from these samples was 90 percent;

(2) in (20), the unknown $\lambda(r)$ had to be estimated by $\hat{\lambda}(r)$.

From (1) and (2), it follows that if $\hat{\lambda}(r)$ underestimates $\lambda(r)$ by more than 10 percent, $\hat{p}$ will exceed 100 percent.

There is, however, a way of avoiding this problem. Since we are practically interested here in a lower limit only, which will enable us to say that, based on a given sample of data and a level of significance α, the true systems relevance p is better than some lower limit $\underline{\hat{p}}$, we can use a different method, to determine a one-sided confidence interval about p. This would correspond to a probability statement

$$P[p > \underline{\hat{p}}] = 1 - 2\alpha = 95 \text{ percent} \qquad (21)$$

which is, for $\alpha = 0.025$, equivalent to saying: the probability that precision p is greater than a lower limit $\underline{\hat{p}}$ is equal to 95 percent.

Lower limits for p determined in this fashion and the corresponding values of $\hat{p} = \Sigma x_i / \Sigma n_i$ are given in table I, for each subset of retrieval runs with a given value of r, for r=1, 2,...., 18. For r > 18, the samples were too small to allow the determination of meaningful limits.

Table I. One-Sided Confidence Intervals on Precision p, Estimated from Samples of Data with Constant r

r	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$\hat{p}$(%)	77	63	100	78	71	87	92	90	92	88	83	75	92	89	100	90	100	100
$\underline{\hat{p}}$(%)	59	52	79	61	54	73	74	78	74	63	55	58	71	76	75	69	77	73

An example may serve to explain how the data from table I can be used to interpret our test results: From the group of retrieval runs performed for all document-based questions with r = 8, we estimate the precision to be 90 percent. Based on this sample of data, we are also confident* to say that the true precision p is better than 78 percent. Going to the next sample with r = 9, we would estimate p to be 92 percent, with a lower limit of 74 percent. The smaller sample size for r = 9 causes a decrease in the lower limit for p, since k, the number of trials, appears in the denominator of the formula (19) for $\hat{\rho}$. Results for the other samples should be interpreted in the same way. For the majority of samples, the estimate for p is better than 80 percent, and the lower limit better than 70 percent. This result should allow an evaluator of ABC systems performance to be confident that the precision of the system under test conditions is better than 70 percent. Unfortunately, the lower limits on p had to be determined for small samples from subsets of data with constant r, since the quantity $\hat{\lambda}(r)$ that depends on r, was involved in the calculating. If precision estimated from all test data ($\hat{\rho}$ = 85.7 percent) would have served as a basis for a determination of p, the lower limit would have been better, maybe close to 80 percent.

C. SUMMARY

We can summarize the results of our discussion of the errors involved in the estimation of relevance and recall from our test data in the following statements:

(1) Recall, defined as the expected value of the average recall ratio, is high when estimated from samples of data with a low value of r. For example, retrieval runs (about 80 runs in the test) for which there was only one relevant document in the collection (r = 1) yielded an estimated recall of more than 80 percent.

When estimated from samples with somewhat larger values of r, say r = 2-5, recall decreases to 33 percent; the corresponding lower limits decrease from about 60 to 20 percent. At still higher values of r up to r = 10, recall shows a further decrease,—though at a much slower rate — to about 17 percent with a lower limit of 15 percent. The reader should remember that the model estimates recall simply as the average recall ratio observed; therefore, the estimates given here represent data observed in the test. It should also be mentioned at this point (see reference 3 for detailed discussion) that the actual average recall ratios observed

*The degree of this subjective confidence can be specified by relating it to the numerical value of the error probability 2α = 5 percent, in (21).

in the test were only a few percent short of the optimal recall ratios obtainable for the observed number of documents retrieved.*

(2) The model allows systems relevance p to be estimated effectively both from all data as well as from subsets with constant r; to a first approximation, the estimates of p obtained from subsets of data with constant r appear to be independent of r. For a level of significance of α = 2.5 percent, the lower limits on p, as determined from a one-sided interval, are greater than 50 percent for all samples with a constant r, r = 1 to 18, the majority being greater than 70 percent. On the basis of all test data (fig. 1), we estimate the ABC systems relevance as $\hat{p}$=85.7 percent. Based on subsets of test data with constant r(r=1,2,...18), we find $63 < \hat{p} < 100$ percent, and for the majority of samples, $\hat{p} > 80$ percent.

Evidently, the model is successful in as much as a basic assumption, the existence of a constant precision parameter p, is confirmed by the test data.

The soundness of the model is further corroborated by the fact that the second assumption, the equivalence of the retrieval process under test conditions with a Poisson process, leads to probability laws for the variables x, y and n that are closely approximated by the observed data.

---

* If n documents are retrieved in a particular run, the optimal recall ratio obtainable is evidently n/r, e.g. all documents retrieved are relevant.

APPENDIX A.—Rationale for the Assumption of a Poisson Distribution for X

(1) The Poisson Process and the Poisson Distribution

This development follows closely that given in Cox and Miller (pp 146 ff).

A Poisson process is a point process on the real axis. We let $N(t,t+\Delta t)$ be the number of these events that occur in the interval between t and $t+\Delta t$. The length, Δt, of the interval is assumed to be small. Let ρ be a positive constant. We further assume

$$\text{Prob}\left[N(t,\ t+\Delta t) = 0\right] = 1 - \rho\Delta t + o(\Delta t)$$

$$\text{Prob}\left[N(t,\ t+\Delta t) = 1\right] = \rho\ \Delta t + o(\Delta t)$$

so that

$$\text{Prob}\left[N(t,\ t+\Delta t) \geq 2\right] = o(\Delta t)$$

also $N(0,\ t)$ and $N(t,\ t+\Delta t)$ are independent. $o(\Delta t)$ is a number smaller than Δt. For most practical purposes we may assume it to be zero. Roughly speaking we have assumed: the probability of one event occurring in a small interval Δt is proportional to the length of the interval; the probability of more than one event occurring in the interval is zero; on the average ρ events will occur per unit of measurement; the occurrence of an event in one interval does not affect the occurrence in another disjoint interval. This is the Poisson process.

Let $N(t) = N(0,\ t)$ = number of events that occur in the interval $(0,\ t)$ and let $p_i(t) = \text{Prob}\ [N(t) = i]$, $i = 0, 1, \ldots$

$$\begin{aligned} p_i(t + \Delta t) &= \text{Prob}\ \{N(t+\Delta t) = i\} = \\ &\quad \text{Prob}\ \{N(t) = i \quad \text{and } N(t,\ t+\Delta t) = 0\} \\ &+ \text{Prob}\ \{N(t) = i-1 \text{ and } N(t,\ t+\Delta t) = 1\} \\ &+ \sum_{k=2}^{i} \text{Prob}\ \{N(t) = i-k \text{ and } N(t,\ t+\Delta t) = k\} \\ &= p_i(t)(1-\rho\Delta t+ o(\Delta t)) + p_{i-1}(t)(\rho\Delta t + o(\Delta t)) \\ &+ \sum_{k=2}^{i} p_{i-k}(t)\ o(\Delta t) \end{aligned}$$

By the independence of "N" on nonoverlapping increments, thus letting $p_{-1}(t)$ be zero, we have the formula

$p_i(t + \Delta t) = p_i(t)(1-\rho\Delta t) + p_{i-1}(t)\,\Delta t\,\Delta t + o(\Delta t)$

which yields the differential equations

$$p_i'(t) = -\rho p_i(t) + \rho p_{i-1}(t) \text{ and the boundary conditions}$$

$$p_o(0) = 1,\ p_i(0) = 0 \quad i = 1, 2, \ldots$$

Consider now the generating function

$$G(Z, t) = \sum_{i=0}^{\infty} p_i(t)\, Z^i$$

$$\frac{\partial G}{\partial t} = \sum_{i=0}^{\infty} p_i'(t)\, Z^i = \sum_{i=0}^{\infty} \left[-\rho p_i(t) + \rho p_{i-1}(t)\right] Z^i$$

$$= -\rho G(Z, t) + \rho Z\, G(Z, t) = \rho(Z-1)\, G(Z, t)$$

then

$$G(Z,t) = A(Z)e^{-\rho t + \rho tZ}$$

Since $G(Z,0) = \sum p_i(0)Z^i = 1$

by the boundry conditions, $A(Z) = 1$

and

$$\text{Prob}\left[N(t) = i\right] = p_i(t) = \frac{d}{dZ}\ G(Z, t)\Big|_{Z=0}$$

$$= \frac{e^{-\rho t}(\rho t)^i}{i!}$$

Hence N(t) = number of events that occur in the interval (0, t) has a Poisson distribution with parameter ρt

(2) <u>The Retrieval Process</u>

Through the arrangement of document descriptions in the ABC index, any given question will guide the searcher to certain subsets of descriptions.

In a first approximation, reference is provided from the keywords of the question to the corresponding cluster of descriptions permutated under those keywords or under synonymous terms. These subsets of descriptions may contain, in typical cases, between 20 and about 100 items, with a few descriptions pertaining to relevant documents interspersed. The searcher will then scan through the subset of descriptions and select those that seem to pertain to relevant documents. After the search is complete, the documents pertaining to the selected descriptions are evaluated with regard to their relevance to the inquiry (ref 3, page 6). The total set of descriptions selected is then divided into a subset pertaining to relevant documents and a subset pertaining to nonrelevant documents.

(3) Retrieval as a Poisson Process

A correspondence between the mathematical process outlined in (1) and the physical process outlined in (2) can now be established. We shall first discuss the retrieval of relevant documents. The subset of descriptions scanned corresponds to the real axis. The selection of a relevant description* is a point event on this axis and $N(t,t+\Delta t)$ the number of descriptions selected in an interval Δt.

Our Δt consists of one description. Hence $N(t,t+\Delta t)$ can be either zero or one. If we now say that there is some rate ρ_1 at which the relevant descriptions will be selected, the correspondence between the two processes is complete. Finally it now follows that the number of relevant documents retrieved from this subset has a Poisson distribution with parameter $\rho_1 d_1$ where d_1 is the number of documents in the subset.

Similarly, there is a process with the same type of factors operating for the retrieval of nonrelevant documents. The number of those that are selected has a Poisson distribution with parameter $\rho_2 d_2$.

---

*The term "relevant description" stands for "description pertaining to a relevant document."

APPENDIX B.—Derivation of $f(x|n)$.

Let x, y, and n be defined as on page 7. Assume x and y are independent and Poisson distributed:

$$g_1(x, \mu) = \frac{e^{-\mu}\mu^x}{x!}$$

$$g_2(y, \nu) = \frac{e^{-\nu}\nu^y}{y!}$$

Let $Q_x(t)$ and $Q_y(t)$ denote the characteristic functions of x and y.

Then

$$Q_{x+y=n}(t) = Q_x(t)\, Q_y(t) = e^{-\mu(1+e^{it})} e^{-\nu(1+e^{it})} = e^{-(\mu+\nu)(1+e^{it})}$$

Letting $\nu + \mu = \lambda$, we have

$$Q_n(t) = e^{-\lambda(1+e^{it})}$$

and the pmf of n, say g(n), is obtained as:

$$g(n) = \frac{e^{-\lambda}\lambda^n}{n!} .$$

Now let h(x, n) denote the joint pmf of x and n

Then

$$f(x|n) = \frac{h(x,n)}{g(n)}$$

$$= \frac{h(x, y = n - x)}{g(n)}$$

$$= \frac{\dfrac{e^{-\mu}\mu^x}{x!} \cdot \dfrac{e^{-\nu}\nu^{(n-x)}}{(n-x)!}}{\dfrac{e^{-\lambda}\lambda^n}{n!}}$$

$$= \frac{n!}{(n-x)!\ x!} \left(\frac{\mu}{\lambda}\right)^x \left(\frac{\upsilon}{\lambda}\right)^{(n-x)}$$

Letting $p = \frac{\mu}{\lambda}$, and $1-p = q = \frac{\upsilon}{\lambda}$,

we finally obtain the pmf of x, given n:

$$f(x \mid n) = \binom{n}{x} p^x q^{(n-x)}.$$

APPENDIX C.—Derivation of the Maximum Likelihood Estimates for p and $\lambda(r)$ Based on a Series of Observations of (X, n).

Suppose we observe the outcomes of k trials

$$Z_1 = \frac{X_1}{n_1},\ Z_2 = \frac{X_2}{n_2},\ \ldots\ .\ Z_k = \frac{X_k}{n_k}$$

The advantage of the maximum likelihood method is now that the distribution of the Z_i need not be known. All we need is the joint pmf for all observed pairs (X_i, n_i), $i = 1 \ldots k$. Since (6) holds for each pair (X_i, n_i), we have for the joint pmf for all pairs

$$H = \prod_{i=1}^{k} h(X_i, n_i, p) = \left\{\prod \binom{n_i}{X_i}\right\} p^{\Sigma X_i} q^{(\Sigma n_i - \Sigma X_i)} e^{-k\lambda(r)} \lambda^{\Sigma n_i} \Big/ \prod_{i=1}^{k} (n_i!)$$

Besides p, this distribution contains $\lambda(r)$ as parameter. Next, we define the likelihood function $L \equiv \ln H$:

$$L = \ln \prod \binom{n_i}{X_i} + \Sigma X_i \ln p + (\Sigma n_i - \Sigma X_i) \ln q - k\lambda(r) + \Sigma n_i \ln \lambda(r) - \ln \left\{\prod (n_i!)\right\}$$

Estimates for the parameters p and $\lambda(r)$ can now be determined by maximizing L with regard to these parameters. We obtain $\hat{p}$ by solving $\frac{\partial L}{\partial p} = 0$ for p:

$$\frac{\partial L}{\partial p} = \frac{\Sigma X_i}{p} - \frac{(\Sigma n_i - \Sigma X_i)}{1 - p} = 0$$

$$\hat{p} = \frac{\sum_{i=1}^{k} X_i}{\sum_{i=1}^{k} n_i}$$

$\hat{\lambda}(r)$ is obtained in a similar manner by solving $\frac{\partial L}{\partial \lambda(r)} = 0$ for $\lambda(r)$:

$$\frac{\partial L}{\partial \lambda(r)} = -k + \frac{\Sigma n_i}{\lambda(r)} = 0$$

$$\hat{\lambda}(r) = \frac{\sum_{i=1}^{k} n_i}{k}$$

Appendix D.—Derivation of the Recall Parameter $\rho(r)$

(1) For a Single Observation of (x, r).

Let $\gamma = \frac{x}{r}$ and $\rho_1(r) = E(\gamma)$.

Substituting γr for x in (6) we obtain the joint pmf of γ and n

$$\ell(\gamma, n; p, r) = \binom{n}{\gamma r} p^{\gamma r} q^{n-\gamma r} e^{-\lambda(r)} [\lambda(r)]^n / n!$$

$$\gamma = 0, \frac{1}{r}, \frac{2}{r} \ldots; \quad n = \gamma r, \gamma r + 1, \ldots$$

Summing this pmf over the values of n we obtain the pmf of γ.

$$\upsilon(\gamma, p, r) = \sum_{n=\gamma r}^{\infty} \ell(\gamma, n; p, r) = \frac{p^{\gamma r} e^{-\lambda(r)} \lambda(r)^{\gamma r}}{(\gamma r)!} \sum_{n=\gamma r}^{\infty} \frac{[\lambda(r) q]^{n-\gamma r}}{[n-\gamma r]!}$$

$$= [\lambda(r) p]^{\gamma r} e^{-\lambda(r)} e^{-\lambda(r) q} / (\gamma r)!$$

$$= [\lambda(r) p]^{\gamma r} e^{-\lambda(r) p} / (\gamma r)!$$

$\gamma r = 0, 1, \ldots$ or equivalently

$\gamma = 0, 1/r, 2/r, \ldots$

Thus γr has a Poisson distribution with mean $E[\gamma r] = \lambda(r)p$ and since r is a constant

$$\rho_1(r) = E[\gamma] = \frac{\lambda(r)p}{r}$$

The moment generating function $M_{r\gamma}(t)$ of $r\gamma$ is defined by;

$$M_{r\gamma}(t) = E(e^{r\gamma t}) = \sum_{r\gamma=0}^{\infty} e^{r\gamma t} (\gamma; \rho, r)$$

$$= e^{-\lambda(r)p} \sum (e^t \lambda(r)p)^{r\gamma} / (r\gamma)!$$

$$= e^{-\lambda(r)p}\, e^{e^t \lambda(r)p}$$
$$= e^{-\lambda(r)p[1-e^t]}$$

We will now derive $\rho(r)$ for k subsequent inquiries performed either by k different operators on one question or by one operator on k different questions with constant r.

(2) Derivation of $\rho(r)$ for k Observations of (x,r)

We define the average recall ratio for k runs with constant r by

$$\bar{\gamma} \equiv \frac{1}{k}\sum_{i=1}^{k}\gamma_i = \frac{1}{rk}\sum x_i \text{ and } \rho_2(r) = E[\bar{\gamma}]$$

The moment generating function of $rk\bar{\gamma}$ is

$$M_{rk\bar{\gamma}}(t) = E(e^{rk\bar{\gamma}t}) = E\left[e^{(\Sigma r\gamma_i)t}\right]$$

$$= \prod E\left[e^{(r\gamma_i)(t)}\right] = \prod M_{r\gamma_i}(t)$$

$$= \left[M_{r\gamma}(t)\right]^k = e^{-k\lambda(r)p[1-e^t]}$$

Since the moment generating function is unique, $rk\bar{\gamma}$ is Poisson distributed with mean $k\lambda(r)p$

Hence

$$\rho_2(r) = \frac{p\lambda(r)}{r} = \rho_1(r)$$

and we define $\rho(r) = \rho_1(r) = \rho_2(r) = \frac{p\lambda(r)}{r}$

APPENDIX E.—A Second Look at the Problem

(1) Contingency Tables and Conditional Probabilities

We will now briefly discuss the relationship between the model developed so far and a similar approach* recently suggested, which in turn makes use of the notions and concepts introduced by John A. Swets (ref 8) and R.A. Fairthorne (ref 9).

Swets arranges the important variables of a retrieval subsystem in a 2 x 2 contingency table with attributes R for retrieved and p for pertinence (here synonymous with relevance).

	P: Pertinent	$\overline{p}$: Nonpertinent	Totals:
R: Retrieved	a	b	a + b
$\overline{R}$: Nonretrieved	c	d	c + d
Totals:	a + c	b + d	a + b + c + d

Here, a, b, c, and d denote the frequencies of occurrence of the four conjunctions; e.g., a is the number of pertinent items retrieved, etc.

The following table shows the four possible conjunctions of R, $\overline{R}$, P, $\overline{P}$ that can be derived from Figure 4 in relation to the four basic retrieval situations and their conventional designations:

Conjunctions	Retrieval situation	Conventional designation
(a) $(P \cdot R)$	Pertinent, retrieved	hit
(b) $(\overline{P} \cdot R)$	Nonpertinent, retrieved	false drop
(c) $(P \cdot \overline{R})$	Pertinent, nonretrieved	miss
(d) $(\overline{P} \cdot \overline{R})$	Nonpertinent, nonretrieved	correct rejection

Based on these situations, Swets defines four conditions probabilities:

$Pr_p(R)$ = cond prob. for a pertinent item to be retrieved

$Pr_{\overline{p}}(R)$ = cond prob. for a nonpertinent item to be retrieved

$Pr_p(\overline{R})$ = cond prob. for a pertinent item to be missed

*Arthur D. Little, personal communication

$Pr_{\bar{p}}(\bar{R})$ = cond prob. for a nonpertinent item to be missed

They can be estimated by the following functions of the frequencies of occurrence a, b, c, and d:

(a) $\frac{a}{a+c}$ estimates $Pr_p(R)$

(b) $\frac{b}{b+d}$ estimates $Pr_{\bar{p}}(R)$

(c) $\frac{c}{a+c}$ estimates $Pr_p(\bar{R})$

(d) $\frac{d}{b+d}$ estimates $Pr_{\bar{p}}(\bar{R})$

The conditional probabilities defined by Swets are not exhaustive; additional probabilities are definable based on the same set of four retrieval situations, simply by reversing the sequence of the attributes "P" and "R" to obtain another set of four conditional probabilities, the first of which is $Pr_R(P)$ = "cond. prob. for a retrieved document to be pertinent" is identical with our precision parameter p. In the following section, we will introduce a second probability model which is based on the four conditional probabilities as defined; since the basic parameter of this model is recall "α"* as compared with precision "p" in the first model, we will distinguish the two by denoting them "α-model," or "p-model"; respectively.

(2) <u>The α-Model</u>

Let us now define

$Pr_p(R) \equiv \alpha$ (probability of a hit, recall)

$Pr_{\bar{p}}(R) \equiv 1-\beta$ (probability of a false drop)

$Pr_p(\bar{R}) \equiv 1-\alpha$ (probability of a miss)

$Pr_{\bar{p}}(\bar{R}) \equiv \beta$ (probability for correct rejection),

and let x_i denote the number of relevant documents retrieved (in response to the ith inquiry) out of a collection of N documents where it is known that r_i documents are relevant to the inquiry.

*following A. D. Little notation

Further let y_i denote the number of nonrelevant documents retrieved; hence $n_i = x_i + y_i$ denote the total number of documents retrieved. With these definitions and under the assumption that sampling takes place with replacement, the distribution of x_i is binomial with mean αr_i, the corresponding pmf being

$$f(x_i, \alpha) = \binom{r_i}{x_i} \alpha^{x_i}(1-\alpha)^{r_i - x_i} \qquad x_i = 0, 1, 2, .., r_i$$

$$g(y_i, (1-\beta)) = \binom{N-r_i}{y_i}(1-\beta)^{y_i} \beta^{N-r_i-y_i}; \; y_i = 0,1,2...N-r;$$

For the combined density of $x_i + y_i = n_i$, we have

$$h(n_i) = \sum_{j=0}^{r_i} [f(x_i = j) \cdot g(y_i = (n_i - j))]$$

with mean

$$E[n_i] = \alpha r_i + (N - r_i)(1-\beta)$$

These equations together with the definitions and assumptions listed on p. 32, form the basis for the α-model. It will be shown, however, that the new model subsequently called α-model, will under certain conditions lead to the same results as the p-model, for:

(a) the maximum likelihood estimate for α, the hit probability of the α-model, which corresponds to $\rho(r)$, the recall parameter in the p-model.

(b) the pmf for the observed recall-ratio x/r.

The conditions are essentially those which relate the binomial pmf to the Poisson pmf, i.e. for large N, large r and small α such that $r\alpha$ remains finite, the Poisson pmf closely approximates the binomial.

(3) Relation between α-Model and p-Model

(a) The Maximum Likelihood Estimate of the Hit-Probability α

The joint pmf of responses x_i, i=1, 2, ...k, to k inquiries for each of which there are exactly r relevant documents in the collection, is given by using

$$f(x_1, x_2, \ldots x_k, \alpha) = \prod_{i=1}^{k} f(x_i, \alpha) = \prod_{i=1}^{k} \binom{r}{x_i} \alpha^{x_i}(1-\alpha)^{r-x_i}.$$

$$= \left\{ \prod_{i=1}^{k} \binom{r}{x_i} \right\} \alpha^{\Sigma x_i} (1-\alpha)^{(kr-\Sigma x_i)}$$

The corresponding likelihood function is

$$L = \ln f(x_1, x_2, \ldots x_i, \alpha) = \sum_{i=1}^{k} \ln \binom{r}{x_i} + \sum_{i=1}^{k} x_i \ln \alpha + [kr - \Sigma x_i] \ln(1-\alpha)$$

The maximum likelihood estimate of α, say $\hat{\alpha}$, is that value of α that maximizes L, i.e., the solution to

$$\frac{\partial L}{\partial \alpha} = 0.$$

Hence

$$\frac{\partial L}{\partial \alpha} = \frac{\Sigma x_i}{\alpha} - \frac{(kr - \Sigma x_i)}{1-\alpha} = 0; \quad \text{and} \quad \hat{\alpha} = \frac{\Sigma x_i}{kr}$$

Evidently, $\hat{\alpha} = \hat{\rho}(r)$, the estimate for the recall parameter according to the "p" model.

(b) The probability mass function of the recall ratio x/r* states:

$$f(x) = \binom{r}{x} \alpha^x (1-\alpha)^{r-x} ; \qquad x=0, 1, 2, \ldots r.$$

Introducing $z = \frac{x}{r}$ and replacing x by rz, we get

$$g(z) = \binom{r}{rz} \alpha^{rz} (1-\alpha)^{r-rz}; \quad rz = 0, 1, 2, \ldots r$$

Then, for large r and small α such that $r\alpha$ remains finite, it can be shown that the binomial expression for g(z) reduces to

$$g(z) \simeq \frac{e^{-\xi} \xi^{rz}}{(rz)!} ; \qquad rz = 0, 1, 2 \ldots r$$

when the parameter $\xi = r\alpha$.

In the p-model, we had arrived at the same result except that the parameter was $\xi = \lambda(r)p$. ($\lambda_{(r)}$ and p as previously defined for the p - model). Hence we have

$$r\alpha = \xi = \lambda(r)p,$$

$$\alpha = \frac{\lambda(r)p}{r} = \rho(r).$$

*In this paragraph, function symbols that had been introduced in previous sections are used again with a different meaning as defined here.

REFERENCES

(1) B. Altmann, The Medium-Sized Information Service; Its Automation for Retrieval, HDL, TR-1192, 30 Dec 63 (AD 429 242).

(2) B. Altmann, A Multiple Testing of the ABC Method and the Development of a Second-Generation Model, Part I, Preliminary Discussions of Methodology, Supplement: Computer Programs of the HDL Information Systems, HDL, TR-1295, April 1965 (AD 617 118).

(3) B. Altmann, A Multiple Testing of the ABC Method and the Development of a Second-Generation Model, Part II, HDL TR-1296, Oct 65 (AD 625 924).

(4) R. A. Fisher; Contributions to Mathematical Statistics, J. Wiley Sons, N. Y., 1950.

(5) A. M. Mood; Introduction to the Theory of Statistics. McGraw Hill Book Co. Inc., New York, 1950

(6) G. P. Wadsworth and J. G. Bryan; Introduction to Probability and Random Variables, McGraw Hill Book Co., N. Y., 1960.

(7) Biometrica Tables for Statisticians; Vol I., ed. by F.S. Pearson and H. O. Hartley, Cambridge Univ. Press, 1956; pp 75, 130.

(8) J. A. Swets; Information Retrieval Systems, Science 141, 1963, pp 245-250.

(9) R. A. Fairthorne; Basic Parameters of Retrieval Tests. ADI - Proceed, 1964, Vol I, p. 343.

Selected Papers and Reports on Modeling Techniques, Especially as Applied to the Development, Description and Performance Testing of IR Systems (with Short Annotations)

(10) J. Verhoeff et al; Mathematical Models in Systems Design for Information Retrieval. Western Reserve University, School of Library Science, May 1961, Contract No. AF 49(638)-357. Contains a description of a probabilistic model for ISR-Systems.

(11) Stephen Pollock; The Normalized "Sliding" Ratio Measure. Arthur D. Little, Inc., Tech. Note CaCl 19 Jul 65. Presents a Mathematical Model of an IR-System, for which generalized recall and relevance are derived as performance measures.

(12) W. Goffman, V. A. Newill, Methodology for Test and Evaluation of Information Retrieval Systems. Western Reserve Univ., School of Library Science, Report No. CSR-TRI, July 1964. (AD 614 005). Sophisticated performance evaluation measures are derived for an idealized IR-system which is described using mathematical and set theoretical concepts and notation.

(13) A. Trachtenberg et al; An Investigation of the Techniques and Concepts of Information Retrieval. ITT Internat. Electric Corp., Tech Rep No. P-AA-TR-(0031), Contr No. DA-36-039-SC-90787. Within the objectives of developing a theory of information retrieval, a preliminary mathematical-probabilistic model of an ISR-system is presented; quantitative measures of relatedness are established.

(14) W. Karush, On Mathematical Modeling and Research in Systems. System Development Corp., SP-1039, Nov 1962. General treatise on systems modeling, especially on mathematical models.

(15) C. P. Bourne et al.; Requirements, Criteria and Measures of Performance of ISR-Systems. Stanford Research Inst., Dec 1961. SRI Proj. No. 3741 (AD 270 942). Contains description of a general functional model (flow-chart type) of an IR-system; discusses future research in modeling for performance evaluation.

(16) R. P. Heckman; A Method for Investigating the Behavior of Attributes which Belong to ISR Systems. Georgia Inst. of Tech., Aug. 1965. Master's Thesis. (AD 624 658). Mathematical statistical model for ISR-system is developed based on functional relationships between ISR-system attributes. Statistical analysis is performed on data from a representative sample of ISR-systems.

(17) J. M. Hoffmann; Experimental Design for Measuring the Intra- and Inter-Group Consistency of Human Judgement of Relevance. Georgia Inst. of Tech., Aug. 1965, Master's Thesis. (AD 620 342). Demonstrates the applicability of statistical methods to the evaluation of tests of relevance assessment consistency.

(18) C. R. Blunt; An Information Retrieval System Model. HRB-Singer Inc., October 1965. Rep. No. 352. 14-R-1. (AD 623 590) A computer simulation model for the performance evaluation of intelligence-type IR-systems.

(19) D. F. Votaw, Jr.; Statistical Science and Information Technology; in: Proceed. of the Second Congress of the Information System Sciences. Spartan Books Inc., Washington, D. C. 1965. Survey of the application of various statistical methods and tools to problems in information technology, i.e. operations research, management science, systems analysis, etc.

(20) D. R. Swanson; On Indexing Depth and Retrieval Effectiveness; in: Proceed. of the Second Congress on the Information System Sciences. Spartan Books Inc., Washington, D. C. 1965. Mathematical statistical model for the evaluation of IR-system performance. Derivation of an analytical expression for the relation of recall and relevance to indexing depth. Application to Cranfield results.

REFERENCES (Continued)

(21) J. F. Rial; Results of Document Retrieval Experiment on Mextrix Searching. Mitre Corp., April 1964. TM-03989. Presents an abstract algebraic model for document retrieval.

(22) C. R. Conger; The Simulation and Evaluation of IR-Systems. HRB-Singer Inc., April 1965. Rep. No. 352-R-17 (AD 464 619) A simulation model to study response time aspects of computer-based IR-systems; practical applicability limited.

(23) H. Borko, The Conceptual Foundations of Information Systems. SDC, May 1965, Rep. No. SP-2057 (AD 615 718). A conceptual model of advanced ISR-systems; discusses prospects of automated indexing and abstracting.

(24) J. Marschak, K. Miyasawa; Economic Comparability of Information Systems. UCLA, Western Management Science Inst., July 1965. Working Paper No. 85 (AD 619767). Models based on decision and utility theory for the comparative evaluation of information systems.

(25) D. J. Hillman; Study of Theories and Models of Information Storage and Retrieval. (Series Title). Lehigh Univ., 1962-1965. Research supported by the National Science Foundation. Investigation on various related problems, including a Boolean algebra model for an IR-system, a graph-theoretical treatment of relatedness of documents and numerous other topics relevant to modeling.

(26) R. Jernigan, A. G. Dale; Set-Theoretic Models for Classification and Retrieval. Univ. of Texas Linguistics Res. Center, Nov. 1964, LRC64-WTM-5. Models based on notions and axioms from lattice theory and topology are suggested for the analysis of ISR-systems.

(27) G. A. Markel; Toward a General Methodology for Systems Evaluation. HRB-Singer, Inc., July 1965. Rep. No 352-R-13 (AD 619 373). Contains an annotated bibliography of literature on systems modeling and simulation. See also; Rep. No. 352.14-R-2, April 66.

(28) A. J. Sailer; Linear Prediction Models for a Mechanized Information System, University of Pittsburgh, Master's Thesis 1966 (AD 481444) Cont AF 33(608) - 1768.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
Harry Diamond Laboratories
Washington, D. C. 20438

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP

3. REPORT TITLE
MULTIPLE TEST OF ABC METHOD PART III—MATHEMATICAL MODEL

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*
W. Menden

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
May 1967	44	28

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. DA-1L013001A91A	TR-1334
c. AMCMS Code: 5016.11.84400	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d. HDL Proj. No. 01200	

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	ARO

13. ABSTRACT

The report suggests a method of constructing a mathematical model for the first test of the ABC Storage and Retrieval Systems and calculates 95-percent confidence intervals for relevance and recall values.

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Retrieval systems						
Evaluation						
Operations analysis						
Statistical analysis						
Modeling						
Documentation						
Technical information centers						